İktisat Politikası Araştırmaları Dergisi -
Journal of Economic Policy Researches
Cilt/Volume: 9, Sayı/Issue: 2, 2022
E-ISSN: 2148-3876

İSTANBUL UNIVERSITY PRESS

RESEARCH ARTICLE / ARAŞTIRMA MAKALESİ

Can Google Search Data Improve the Unemployment Rate Forecasting Model? An Empirical Analysis for Turkey

Google Arama Verileri İşsizlik Oranı Tahmin Modelini İyileştirebilir mi? Türkiye için Ampirik Bir Analiz

Gülşah ŞENTÜRK[1]

ABSTRACT

Today, data accumulated during internet use have become an important source of information for people's behaviour, issues, and needs, and due to real-time data acquisition, Google search data have become a focal point for researchers. As a result, it has been become more common to use GT data, which have been included in forecasting models for many economic indicators, including unemployment rate forecasting. Therefore, this study aims to determine whether including Google search data in forecasting models can improve the model's performance in forecasting the unemployment rate in Turkey. In this context, out-of-sample forecasting was performed in this study using seasonally adjusted monthly unemployment rates for the period between January 2005 and August 2020 and monthly GT data about the topic of unemployment insurance. In addition, the forecasting performance of ARIMA and ARIMAX methods were compared.

Keywords: Google trends, Unemployment rate, Time-series model, Forecasting, ARIMA

Jel Code: C53, E24, E37

ÖZ

İnternet kullanımı esnasında depolanan verilerin insan davranışları, sorunları ve ihtiyaçları için önemli bir bilgi kaynağı haline geldiği günümüzde, Google arama verileri gerçek zamanlı olarak elde edilmesi nedeniyle araştırmacıların odağı haline gelmektedir. Pek çok ekonomik gösterge için tahmin modellerine dahil edilmeye başlanan Google Trends verilerinin işsizlik oranı tahmininde de kullanılması giderek yaygınlaşmaktadır. Bu çalışma, Türkiye'de işsizlik oranının tahmin

DOI: 10.26650/JEPR963438

[1] Res. Assist., Istanbul University, Faculty of Economics, Istanbul , Turkey

ORCID: G.Ş. 0000-0002-4252-4772

Corresponding author/Sorumlu yazar:
Gülşah ŞENTÜRK,
Istanbul University, Faculty of Economics, Istanbul , Turkey
E-mail/E-posta:
gulsahsenturk@istanbul.edu.tr

Submitted/Başvuru: 06.07.2021
Revision Requested/Revizyon Talebi: 04.11.2021
Last Revision Received/Son Revizyon: 05.11.2021
Accepted/Kabul: 16.11.2021

Citation/Atıf: Senturk, A. (2021). Can Google Search Data Improve the Unemployment Rate Forecasting Model? An Empirical Analysis for Turkey. *İktisat Politikası Araştırmaları Dergisi - Journal of Economic Policy Researches, 9*(2), 229-244. https://doi.org/10.26650/JEPR963438

edilmesinde Google arama verilerinin öngörü modeline dahil edilmesinin modelin öngörü yeteneğini iyileştirip iyileştirmediğini araştırmaktadır. Ocak 2005'ten Ağustos 2020'ye kadar olan dönem için mevsimsellikten arındırılmış aylık işsizlik oranları ile işsizlik sigortası konusuna dair aylık Google Trends verileri ele alınarak öngörü modeli oluşturulmaktadır. ARIMA ve ARIMAX yöntemleri aracılığıyla yapılan tahminlerin öngörü performansı kıyaslanmaktadır.

Anahtar Kelimeler: Google trendler, İşsizlik oranı, Zaman serisi modelleri, Gelecek tahmini
Jel Code: C53, E24, E37

1. Introduction

Traditional means of obtaining information can be slow compared to the fast changes in the economy. On the other hand, the internet is a source of information, also called "big data", that can be accessed in real-time. Even though the use of big data analyses is not yet widespread in the field of economics, various platforms that present big data by filtering it in specific ways constitute a remarkable resource for economists. The most common example is the Google Trend (GT) Data constituted by the Google search volume data. GT Data exist for different keywords and topics in many different categories and present scaled search volumes after normalizing them, provided that there is a sufficient number of searches.

As the data can be obtained instantaneously for studies in numerous fields, Google search volumes provide a significant advantage. In addition, GT Data constitute real-time data and can be informative for various economic indicators. Despite the lack of equal opportunities in terms of internet access, today, about 51% of the global population and 74% of the population in Turkey have access to the internet, according to World Bank data (2019). These rates show that internet data concern a significant portion of society and contain meaningful information regarding economic and social indicators, compared to extensive surveys conducted to obtain information about these indicators.

Today, online platforms are commonly used for job searches and job advertisements, while other means are less frequently preferred due to increased opportunities to find jobs in suitable positions and reach people employed online. Google, one of the search engines that people use to access all kinds of information they need at any time of the day, is a platform where people search for these job advertisements and obtain information about unemployment benefits. It is believed that using GT Data instead of unemployment rates calculated based on workforce surveys can make it possible to avoid the prejudicial results of survey management. Another advantage of using GT Data over workforce surveys in calculating unemployment rates is the opportunity for instant access (Baker & Fradkin, 2013).

It is becoming more common in the literature to use Google data to forecast economic indicators and analyse specific trends. In particular, the literature on forecasting unemployment rates has expanded with the increase in studies for many countries in recent years. Unemployment rate forecasting has been done for many developing and developed countries, including Turkey, using GT Data by employing various methods. However, there is a lack of consensus regarding data organization and specification for GT Data. Therefore, certain matters need to be considered when using GT Data. Google provides detailed information regarding calculation methods and scopes of these data. As there are no studies in the literature conducted using data relating to the categorization of seasonally adjusted monthly unemployment rates and GTD as "topics" for Turkey, this study aims to fill the said gap in the literature in this regard and to shed light on future studies by explaining necessary information on how GT Data are to be effectively used.

This study makes use of GT Data in order to improve the standard unemployment forecasting models. The ARIMA comparative model was employed for forecasting by using seasonally adjusted unemployment rates related to Turkey. In addition to the unemployment rates, the forecasting made by the ARIMAX model was compared against the forecasting model of comparison. This way, the study aims to investigate if Google search volumes improve the performance of the unemployment rates forecasting model. In the second section of the article, GT Data, which are considered big data, are analysed, and the significant points to consider when using these data are summarised. The third section summarizes the studies conducted with the GT Data. Afterward, information is given regarding the methodology and the scope of the study. In the fifth section, empirical findings are presented. Finally, in the sixth section, the conclusions reached with this study are discussed.

2. Google Trends Data

Google Trends can be explained as a database that provides data obtained from individuals' internet behaviour via the Google search engine after applying certain normalizations. As individuals' internet behaviour becomes more and more important, big data becomes a source of new and extensive information for researchers as it undergoes a different form of categorization and can be used as open access for different geographical locations.

Google Trends provides information regarding the changes in search volumes over time by categorizing "topic" and "search terms" and by enabling geographical filtering for particular keywords[1] by Google. These data can be obtained for a particular geographical area based on users' IP addresses on global, national, and local levels for the desired period (starting from 2004) and scaled in a form normalized for that particular period (Google, 2017).

Data presented via GT are based on two different samples: real-time and non-real-time. While the non-real-time GT Data are from the last seven days, real-time GT Data exist up to the last 36 hours from 2004 onwards and can be obtained for any desired period. However, the period requested for non-real-time GT Data determines the frequency[2] of the data obtained. Therefore, the possibility for different frequencies for these data increases the chances of using them for many different analyses (Google, 2017).

The most significant point relating to GT Data is not combining data for different periods to create a single time series. It has been observed that in some studies in the literature, data obtained for different periods are combined to collect data for desired frequencies (e.g., daily or weekly) as a single time series to be included in analyses. We wish to emphasize that analyses made from a time series created by combining these data of different segments will not yield healthy conclusions. The fundamental reason behind this emphasis is that the GT Data are presented according to time and location after normalization.

GT Data are normalized by dividing the search volume from each selected geographical location at each time point, based on the frequency of the data sample for the keywords, by the total search volume in all periods selected in this geographical location and scaling the result between 0 and 100. Accordingly, GT Data is calculated for each related geographical location and period as;

$$\frac{\text{Search volume in period t}}{\text{Total search volume for all time}}\,x100 \qquad (1)$$

This kind of normalization is to neutralize the disadvantage of obtaining misleading information based on the population of the geographical areas. However, this normalization means that the same total search volume for different periods will differ. Therefore, the time series that combines different periods to achieve the desired frequency contain misleading information. The GT Data used in the analyses in this study are the monthly frequency data related to the period between January 2005 and September 2020 in Turkey.

[1] Keywords with sufficient levels of search volumes are emphasized. If the words to be investigated do not have sufficient levels of search volumes, they are not included in GT Data.

[2] The data are supplied at various frequencies according to time intervals; the frequency for the last 1-4 hours is 1 minute, for the last 24 hours is 2 minutes, for the last 7 days is 1 hour, for the last 30-90 days is daily, for the last 1-5 years is weekly, and for more than 5 years is monthly.

One of the significant features of GT Data is the differentiation of search terms and topics. Search terms only show the results related to the particular search term. However, if internet users misspell a certain word or use different expressions with the same meaning, these are not included in the search terms. This way, when a word or a phrase is selected in the "topic" categories instead of "search terms" when using GT Data, other words that also give the word and the same search results will be included in the data. Thus, it is believed that using GT Data will produce more extensive information. Considering these points, the phrase "unemployment insurance" is selected as a *topic*, aiming to obtain more extensive information from the data.

3. Literature Review

GT Data is used in studies in a lot of different fields, from flu outbreaks in the field of medicine (Ginsberg, Mohebbi, Patel, Brammer, Smolinski & Brilliant, 2008), exchange rate movements in economics and finance (Goddard, Kita, & Wang, 2015; Smith, 2012), stock markets (Da, Engelberg & Gao, 2011; Aouadi, Arouri & Teulon, 2013; Mondria, Wu & Zhang, 2010; Hamid & Heiden, 2015; Bui & Nguyen, 2019; Huang, Rojas & Convery, 2020), housing prices (Beracha & Wintoki, 2013), private consumption spending (Kholodilin, Podstawski & Siliverstovs, 2010), tourist movements (Bangwayo-Skeete & Skeete, 2015), prices of gold and oil (Han, Lv, & Yin, 2017; Jain & Biswal, 2019), and uncertainty (Donadelli & Gerotto, 2019; Castelnuovo & Tran, 2017), to consumer confidence (Niesert, Oorschot, Veldhuisen, Brons & Lange, 2019). With the forecasting area of macroeconomic indicators expanding with GT Data, several studies on unemployment forecasting are of significance in the literature of analysis that uses Google data as the flagbearer.

Baker and Fradkin (2013) found a strong correlation with the non-linear OLS method for the period 2006-2011 using GT Data, comparing the criteria of job search activity for the term "job" in Texas. Choi and Varian (2009) stated that the autoregressive model, which incorporates the categories "jobs" and "welfare and unemployment" in the USA, and unemployment benefit estimations, particularly turning point estimations, are more effective. D'Amuri and Marcucci (2009), on the other hand, determined that the Google Index they constructed provides a more accurate estimate of unemployment in the USA than the use of unemployment assistance request data. Fondeur and F. Karamé (2013) presented findings indicating that GT Data had better performance in forecasting youth unemployment in France than for other age groups. Barreira, Godinho, and Melo (2013) came to weaker conclusions in studies they conducted for two different areas: forecasting car sales and unemployment rates for Spain, Portugal, France, and Italy by building an autoregressive model for each particular country when compared to other studies. In the causality analysis conducted by Askitas and Zimmerman (2019) regarding the seasonal and seasonally adjusted unemployment rate in monthly frequency for Germany, they found strong correlation in terms of "unemployment bureau[3] or unemployment agency[4]", "unemployment rate[5]", and "personnel consultant[6]" in addition to using GT Data for the most popular job search engines in Germany. D'Amuri (2009b), who used the term "job offers" to forecast the unemployment rate in Italy with GT Data, and Bughin (2011), who included inflation rates along with GT Data with regards to worker's unions in calculating the unemployment rate in Belgium, concluded that Google searches could be used to forecast unemployment. Investigating the relationship between unemployment rates in the Visegrad Countries, namely Czechia, Hungary, Poland, and Slovakia through GT Data, Pavlicek and Kristoufek (2015) established that there was a high correlation between search terms and unemployment rate and that GT Data were beneficial for Czechia and Hungary, while they did not show a strong forecasting performance for Slovakia and Poland.

Two studies in Turkey have used GT Data for unemployment forecasting. Chadwick and Şengül (2012), who conducted the first study in Turkey that forecast unemployment using Google data, developed a forecasting model by classifying weekly GT Data of alternative terms that can be used to search for jobs on the internet using a principal components analysis. This study examined the monthly non-agricultural unemployment rate from 2005 to 2011 and determined that the forecasting model that includes GT Data performed better in forecasting the unemployment rate. In the other study conducted for Turkey, Bolivar, Ortiz and Rodrigo (2019) built a seasonally adjusted unemployment rate forecasting model for April 2007 and June 2019 with variables such as production index, electricity consumption, capacity utilization rate, and the number of applications for unemployment benefit along with the terms "finding a job", "unemployment pay", and "unemployment insurance" for GT Data, and came to the conclusion that this model yielded better results when compared to other models.

4. Methodology

This study use seasonally adjusted unemployment rates[7] for January 2005 and September 2020 for Turkey and the search volume statistics for the topic of "unemployment insurance"[8] in GT Data[9] . In the analyses, the "*un*" variable represents the seasonally adjusted

[3] Arbeitsamt (GER)

[4] Arbeitsagentur (GER)

[5] Arbeitslosenquote (GER)

[6] Personalberater (GER), Personalberatung (GER)

[7] Unemployment rate dataset were obtained from OECD (2021) database.

[8] The use of "topic" is emphasized. "Topic" and "Search Term" are the two different data types for keywords that are provided by GT data.

[9] GT Data were obtained from "https://trends.google.com", which is a search volume tool provided by Google.

unemployment rate while the "*gtinsure"* variable represents GT search volume statistics for unemployment insurance. ARIMA and ARIMAX models were utilized for forecasting with these data.

The Auto-Regressive Integrated Moving Average (ARIMA) model, a time series forecasting model developed by Box and Jenkins (1976), is a frequently used approach. The ARIMA model aims to analyse stochastic time series with a single variable. The ARIMA (p, d, q) model is a forecasting method used for a single variable and is expressed as follows:

$$\Delta Y_t = \delta + \phi_1 \Delta Y_{t-1} + \phi_2 \Delta Y_{t-2} + \ldots + \phi_p \Delta Y_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} + \varepsilon_t \qquad (1)$$

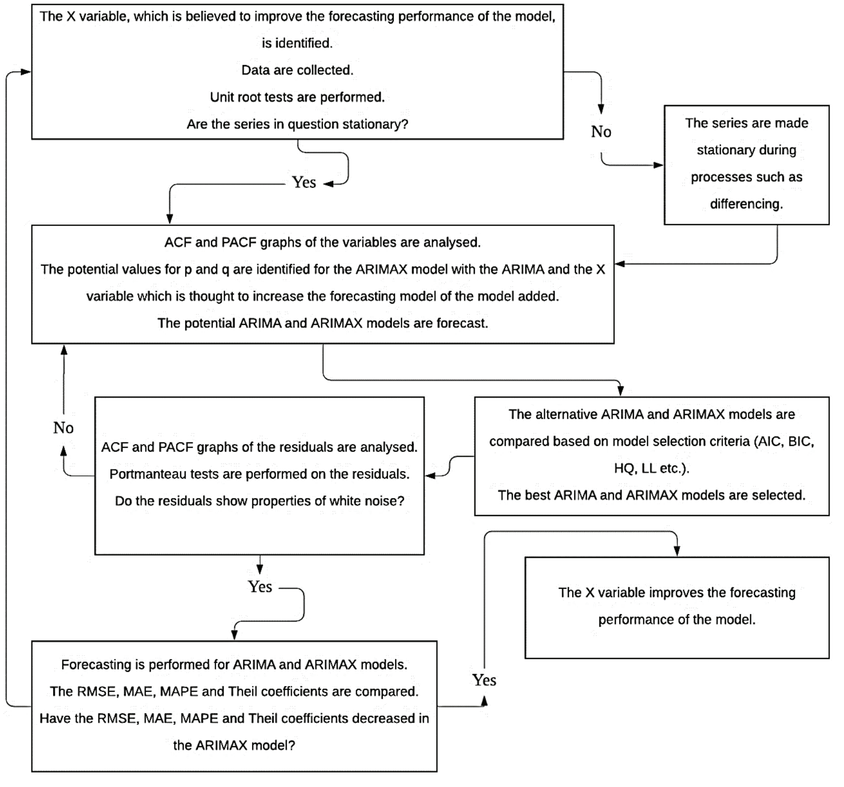
The ARIMA model turns into the ARIMAX model when the independent variable(s) are added to improve the model's forecasting performance. An ARIMAX model with one added independent variable is shown below:

$$\Delta Y_t = \delta + \phi_1 \Delta Y_{t-1} + \phi_2 \Delta Y_{t-2} + \ldots + \phi_p \Delta Y_{t-p} + \beta \Delta X_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} + \varepsilon_t \qquad (2)$$

In the ARIMA and ARIMAX models, a correlogram analysis is performed based on Box-Jenkins methodology, and the appropriate model is selected among the alternatives using the principle of parsimony. Information criteria such as Akaike Information Criterion (AIC), Schwarz Criterion (BIC), and Hannan Quinn (HQ) are a guide in model selection. To determine the best model, the model with the minimum value of the information criteria is searched.

To examine the forecast performance of the alternative models, the Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), and Theil coefficients are compared. These metrics show the amount of deviation from the actual values. Thus, it is determined whether the independent variable added to the model increases the forecast performance of the model (Box & Jenkins, 1976). Accordingly, the stages followed for the study's empirical analysis with Box-Jenkins methodology are shown in Figure 1.

Figure 1. Stages of Forecasting Performance Checking with ARIMA and ARIMAX models

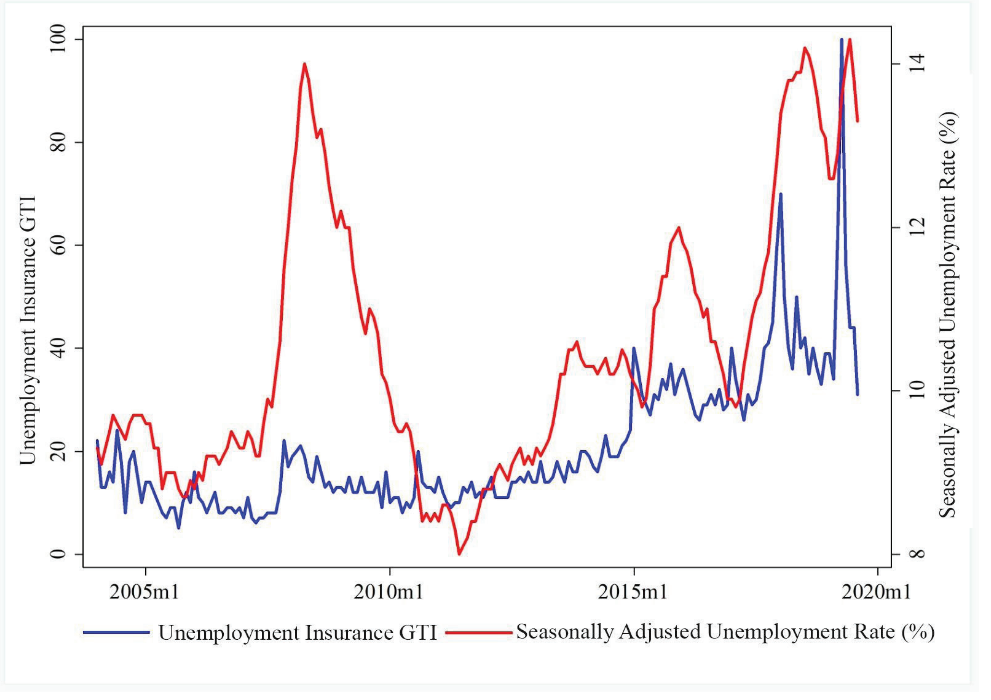


Reference: Sevüktekin, M. & Çınar, M. (2017). *Ekonometrik Zaman Serileri Analizi: EViews Uygulamalı* (Fifth Edition). Bursa: Dora Yayıncılık.

5. Empirical Findings

The seasonally adjusted unemployment rate in Turkey and search volume statistics for unemployment insurance obtained from GT Data are given in Graph 1.

Graph 1. Seasonally Adjusted Unemployment Rate and GTI for Unemployment Insurance



Graph 1 suggests that the statistics of the unemployment rate and that of Google searches related to unemployment insurance show a similar trend. It is possible to see the correlation with the unemployment rate in the graph. The correlation between unemployment statistics and GT Data is determined to be 0.66. The statistics summary showing the "un" variable for the unemployment rate and "gtinsure" variable for the GT Data related to unemployment insurance is given in Table 1.

Table 1: Statistics Summary

	Number of Observations	Average	Standard Deviation	Minimum	Maximum
un	188	10.477	1.632	8	14.3
gtinsure	188	20.632	13.488	5	100

In order to determine whether the GT Data improve the unemployment forecasting model, first, the Box-Jenkins methodology was applied to analyse the stationarity of the variables through unit root tests. Of the unit root tests, ADF and KPSS were employed, and the fundamental hypothesis of the ADF unit root test indicated that the variable was not stationary. In contrast, the fundamental hypothesis of the KPSS indicated that the variable was stationary. The unit root tests were performed when the variables were at their level status based on the model with trend and constants, and root tests were applied on the first differences of the variables based on the model without trend and constants. Table 2 indicates the results of ADF and KPSS unit root tests.

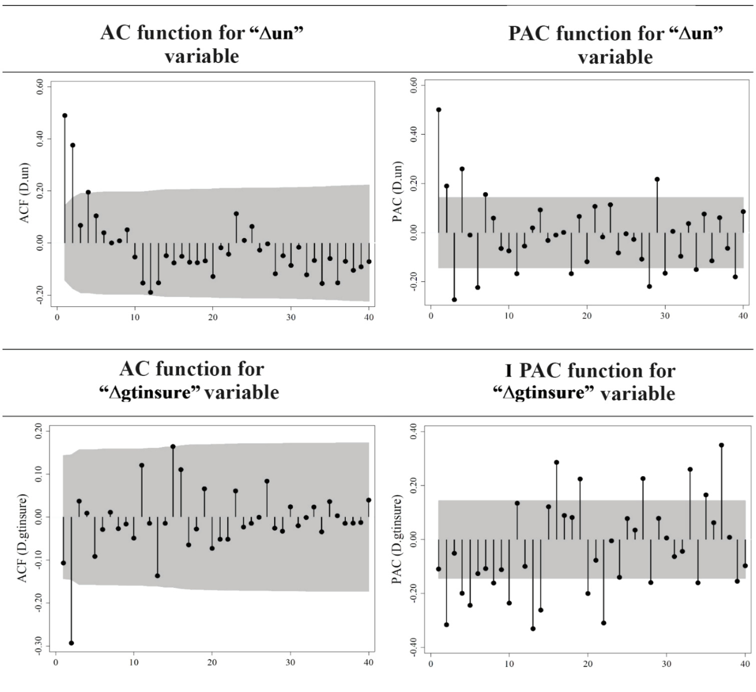
Table 2: Unit Root Test Statistics

	ADF		KPSS	
Variable	Level (Trend and Intercept)	First Difference (None)	Level (Trend and Intercept)	First Difference (None)
un	-1.606	-6.474*	0.181*	0.079
gtinsure	-3.977	-13.845*	0.397*	0.681

* $p<0.05$

According to the unit root tests statistics shown in Table 2, neither variable was stationary at their level and became stationary after the first difference was taken. Therefore, the analysis continued with the first differences, and autocorrelation function (ACF) and partial autocorrelation function (PACF) graphs were analysed. Graph 2 shows the ACF and PACF graphs of the variables.

Graph 2. AC and PAC Functions of the Variables



Note: The confidence interval is 95%.

According to the ACF and PACF graphs shown in Graph 2, p and q values were identified for ARIMA and ARIMAX models based on the delay durations outside of the confidence intervals. Model selection criteria of alternative models for ARIMA and ARIMAX are shown in Table 3.

Table 3: Model selection criteria of alternative models for ARIMA and ARIMAX

ARIMA (p, d, q)	AIC	SIC	HQ	ARIMAX (p, d, q)	AIC	SIC	HQ
(1,1,0)	-0.140	-0.088	-0.119	**(1,1,0)**	-0.175	-0.106	-0.147
(2,1,0)	-0.165	-0.096	-0.137	**(2,1,0)**	-0.192	-0.106	-0.157
(3,1,0)	-0.226	-0.140	-0.191	**(3,1,0)**	-0.245	-0.141	-0.203
(4,1,0)	-0.280*	-0.176*	-0.238*	**(4,1,0)**	-0.298	-0.177	-0.249
(0,1,1)	-0.023	0.028	-0.002	**(0,1,1)**	-0.061	0.007	-0.033
				(0,1,3)	-0.355*	-0.251*	-0.313*

*It is the best model according to the model selection criteria.

Alternative models that give model assumptions based on the identified p and q values were forecast and compared based on the model selection criteria. Based on the difference variables, the ARIMA (4,1,0) and ARIMAX (0,1,3) models are shown as follows:

ARIMA (4, 1, 0) model; $\Delta un_t = \delta + \phi_1 \Delta un_{t-1} + \phi_2 \Delta un_{t-2} + \phi_4 \Delta un_{t-3} + \phi_4 \Delta un_{t-4} + \varepsilon_t$ (3)

ARIMAX (0, 1, 3) model; $\Delta un_t = \delta + \Delta gtinsure_t + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_4 \varepsilon_{t-3}$ (4)

In both models, the autocorrelation functions, partial autocorrelation functions, and Portmanteau tests of the residuals were analysed, and it was shown that the residuals show properties of white noise. Forecasting statistics regarding the ARIMA (4,1,0) and ARIMAX (0,1,3) models are shown in Table 4.

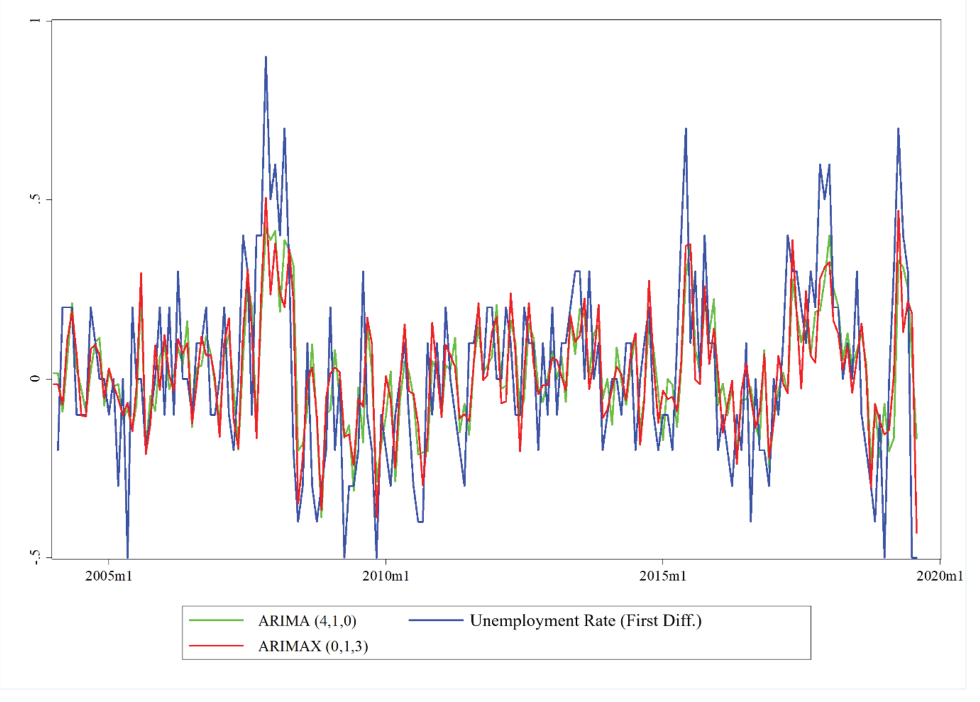
Table 4: Model Statistics

	ARIMA (4,1,0)	ARIMAX (0,1,3)
Constant	0.016-	0,018
AR(1)	0.529*	-
AR(2)	0.216*	-
AR(3)	-0.386*	-
AR(4)	0.258	-
MA(1)	-	0.592*
MA(2)	-	0.558*
MA(3)	-	-0.156*
Δgtinsure	-	0.003*
σ	0.041*	0.038*
R^2	0.367	0.415
Adjusted R^2	0.351	0.399
F statistic	21.04*	25.69*
SSR	7.72	7.14
LL	32.18	39.21
AIC	-0.280	-0.355
BIC	-0.176	-0.252
HQ	-0.238	-0.313

* $p<0.05$

According to the forecasting results of the ARIMA and ARIMAX models, all the variables apart from the constant were significant, and both models are meaningful according to the F-test statistics. While the Adjusted R^2 is 36.7% in the ARIMA model, it increased to 39.9% in the ARIMAX model. When the models are compared according to AIC, BIC, and HQ information criteria, the ARIMAX model is more appropriate. Graph 3 shows the first difference series of the unemployment rate and unemployment rates obtained from the forecast of ARIMA (4,1,0) and ARIMAX (0,1,3) models.

Graph 3. First Difference Series of the Unemployment Rate, Unemployment Rate Forecast Value According to ARIMA (4,1,0) and ARIMAX (0,1,3) Models

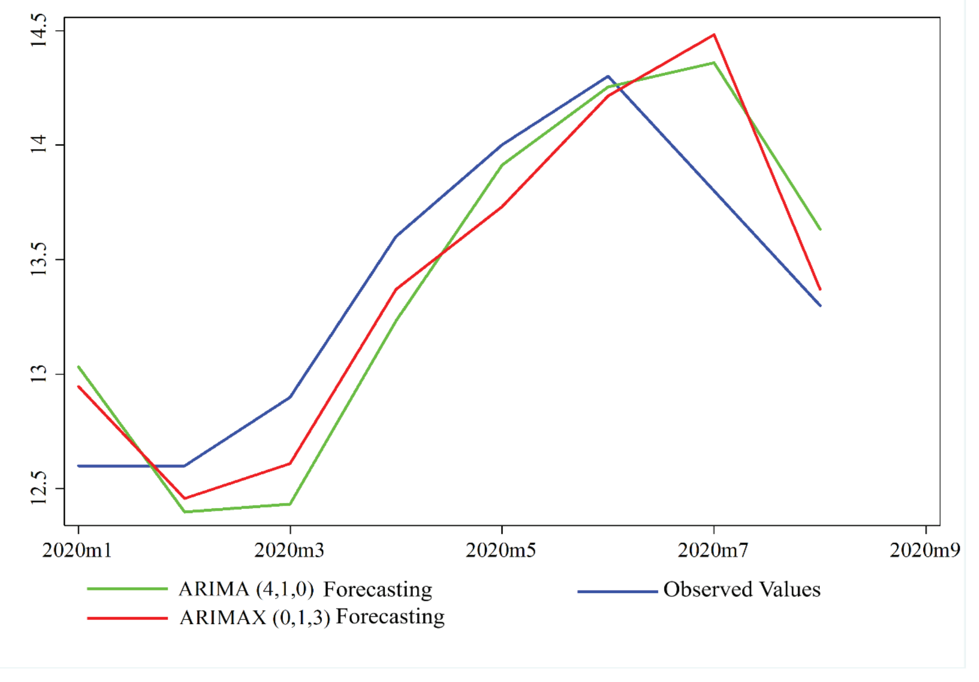


Based on the ARIMA (4,1,0) and ARIMAX (0,1,3) models, out-of-sample forecasting was made. Table 4 and Graph 4 show observed and forecast values for the models.

Table 4: Out-of-Sample Forecasting Values and Observed Values with Regards to the Unemployment Rate

Period	ARIMA (4, 1, 0)	ARIMAX (0, 1, 3)	Observed Values
01-2020	13.0308	12.9446	12.6
02-2020	12.3968	12.4571	12.6
03-2020	12.4333	12.6083	12.9
04-2020	13.2316	13.3701	13.6
05-2020	13.9119	13.7303	14
06-2020	14.2531	14.2158	14.3
07-2020	14.3586	14.4827	13.8
08-2020	13.6331	13.3701	13.3

Graph 4. Out-of-Sample Forecasting and Real Observation Values with Regards to the Unemployment Rate



When the models were compared using the RMSE (Root Mean Square Error) value, the RMSE value in the ARIMA model was 35.5%, while it dropped to 32.8% with the addition of GT Data into the model. Hence, with the addition of Google search volume statistics regarding the topic of unemployment insurance into the standard unemployment rate forecasting model, the forecasting performance of the model improved. The statistics of forecasting performance in Table 5 show that the ARIMAX model (0,1,3) has better forecasting power than the ARIMA (4,1,0) model.

Despite the similarity of the out-of-sample forecasts obtained from the ARIMA and ARIMAX models in Graph 4, the forecast values of the ARIMAX model are shown to be closer to the official unemployment rate data. Table 4 includes forecast performance indicators relating to out-of-sample forecasts of the ARIMA and ARIMAX models such as RMSE, MAE, MAPE, Symmetrical MAPE, Theil coefficient.

Table 5: Performance Indicators of Out-of-Sampling Forecasts

	ARIMA (4,1,0)	ARIMAX (0,1,3)
RMSE	0.355	0328
MAE	0.311	0283
MAPE	2.358	2132
Symmetrical MAPE	2.352	2125
Theil	0.013	0122

When the models were compared using the RMSE (Root Mean Square Error) value, the RMSE value in the ARIMA model was 35.5%, while it dropped to 32.8% with the addition of GT Data into the model. Hence, with the addition of Google search volume statistics regarding the topic of unemployment insurance into the standard unemployment rate forecasting model, the forecasting performance of the model improved. The statistics of forecasting performance in Table 5 show that the ARIMAX model (0,1,3) has better forecasting power than the ARIMA (4,1,0) model.

6. Conclusion

Big data is used in many areas today, and Google Trends has been drawing attention as a data platform that makes the necessary adjustments to the data and is open to everyone. Today, people search for solutions for everything from their daily needs to their most fundamental problems on platforms such as Google search. The data stored by search engines like Google contain society's issues, needs, and thoughts. A significant portion of the Google search volume data, called Google Trends, is open to everyone and very easy to use. This database has made it possible to use a significant source of data like Google to analyse big data frequently and quickly, thanks to the diversity, scope, and accessibility of the data it contains. As a result, studies in numerous fields that focus on data from big data source Google Trends have swiftly progressed.

As people investigate job advertisements or unemployment assistance, mainly on search engines like Google, to find a solution for their concerns and problems, especially regarding their working life, researchers have directed their attention towards using Google data for predictions regarding unemployment rates. Setting off with the hypothesis that Google search volume statistics will improve the unemployment rate forecasting model for the forecast of unemployment rates in Turkey, the models ARIMAX and the standard unemployment forecasting model ARIMA were employed, and their forecasting performances were compared. The findings of the forecasts produced by the models demonstrated that the forecasting performance of the ARIMAX model was higher than that of the ARIMA model. In conclusion, GT Data in forecasting unemployment rates in Turkey can strengthen the forecasting model for the unemployment rate.

When the empirical findings are examined, it is seen that the results are compatible with the other two studies (Chadwick & Şengül, 2012; Bolivar, Ortiz & Rodrigo, 2019) conducted in Turkey. However, unlike the study by Chadwick and Şengül (2012), this study deals with monthly data instead of weekly, and unlike Bolivar, Ortiz, and Rodrigo (2019), it uses search topics instead of search terms. The study by Chadwick and Şengül (2012) emphasized that combining weekly data is an inaccurate data collection method due to the nature of GT Data. In this study, necessary emphases have been made about the use of GT Data, and it is emphasized that the use of search topic data is more comprehensive than the search terms data used by Bolivar, Ortiz, and Rodrigo (2019) in their study.

Peer-review: Externally peer-reviewed.
Conflict of Interest: The author has no conflict of interest to declare.
Grant Support: The author declared that this study has received no financial support.
Acknowledgement: I would like to thank Prof. Dr. Halil Tunalı and Prof. Dr. Ferda Yerdelen for all their support and contributions in the conduct of this study.

References

Aouadi, A., Arouri, M., & Teulon, F. (2013). Investor attention and stock market activity: Evidence from France. *Economic Modelling*, *35*, 674–681.

Askitas, N., & Zimmermann, K. F. (2009). Google econometrics and unemployment forecasting. *Discussion Paper*, 4201.

Baker, S., & Fradkin, A. (2013). The impact of unemployment insurance on job search: Evidence from Google search data. *The Review of Economics and Statistics*, *99*(5), 756–768.

Bangwayo-Skeete, P. F., & Skeete, R. W. (2015). Can Google data improve the forecasting performance of tourist arrivals? Mixed-data sampling approach. *Tourism Management*, *46*, 454–464.

Barreira, N., Godinho, P., & Melo, P. (2013). Nowcasting unemployment rate and new car sales in South-Western Europe with Google Trends. *Netnomics*, *14*, 129–165.

Beracha, E., & Wintoki, M. B. (2013). Forecasting residential real estate price changes from online search activity. *Journal of Real Estate Research*, *3*, 283–312.

Bolivar, Ortiz, & Rodrigo. (2019). Nowcasting Turkish unemployment using real time data from Google. *BBVA Research*.

Box, G., & Jenkins, G. (1976). *Time series analysis, forecasting and control.* Holden-Day: California.

Bughin, J. (2011). Nowcasting the Belgian economy.. *SSRN working paper*.

Bui, V. X., & Nguyen, H. T. (2019). Stock market activity and Google Trends: The case of a developing economy. *Journal of Economics and Development*, *21*(2), 191–212.

Castelnuovo, E., & Tran, T. D. (2017). Google it up! A Google Trends-based uncertainty index for the United States and Australia. *Economics Letters*, *161*, 149–153.

Chadwick, M. G., & Şengül, G. (2012). Nowcasting unemployment rate in Turkey: Let's ask Google. *TCMB Working Paper*, *12*(18).

Choi, H., & Varian, H. (2009). Predicting initial claims for unemployment benefits. *Technical Report*.

D'Amuri, F. (2009b). Predicting unemployment in short samples with internet job search query data. *MPRA Paper*, 18403

D'Amuri, F., & Marcucci, J. (2009). 'Google it!' Forecasting the US uneployment rate with a Google job search index. . *MPRA Paper,* 18248.

Da, Z., Engelberg, J., & Gao, P. (2011).In search of attention. *The Journal of Finance*, 1461–1499.

Donadelli, M., & Gerotto, L. (2019). Non-macro based Google searches, uncertainity, and real economic activity. *Research in International Business and Finance, 48*, 111–142.

Eurostat. (2019). https://ec.europa.eu/eurostat

Fondeur, Y., & F. Karam'e. (2013). Can Google data help predict French youth unemployment? *Economic Modelling, 30*, 117–125.

Ginsberg, J., Mohebbi, H. M., Patel, S. R., Brammer, L., Smolinski, S., & Brilliant, L. (2008). Detecting influenza epidemics using search engine query data. *Nature*, *457*(7232), 1012–1014.

Goddard, J., Kita, A., & Wang, Q. (2015). Investor attention and FX market volatility. *Journal of International Financial Markets*, 79–96.

Google. (2017). https://support.google.com/trends

Google Trends. (2020). https://trends.google.com

Hamid, A., & Heiden, M. (2015). Forecasting volatility with empirical similarity and Google Trends. *Journal of Economic Behavior & Organization*, *117*, 62–81.

Han, L., Lv, Q., & Yin, L. (2017). Can investor attention predict oil prices? *Energy Economics*, *66*, 547–558.

Huang, M. Y., Rojas, R. R., & Convery, P. D. (2020). Forecasting stock market movements using Google Trend searches. *Empirical Economics*, *59*, 2821–2839.

Jain, A., & Biswal, P. C. (2019). Does internet search interest for gold move the gold spot, stock and exchange rate markets? A study from India. *Resources Policy, 61*, 501–507.

Kholodilin, K. A., Podstawski, M., & Siliverstovs, B. (2010). *Do google searches help in nowcasting private consumption? A real-time evidence for the US.* KOF Working Paper.

Mondria, J., Wu, T., & Zhang, Y. (2010). The determinants of international investment and attention allocation: Using internet search query data. *Journal of International Economics*, *82*(1), 85–95.

Niesert, R., Oorschot, J., Veldhuisen, C., Brons, K., & Lange, R. (2019). *Can Google search data help predict macroeconomic series?* Tinbergen Institute Discussion Paper.

Pavlicek, J., & Kristoufek, L. (2015). Nowcasting unemployment rates with Google Searches: Evidence from the Visegrad group countries. *PLOS One*, *10*(5), 1–11.

Sevüktekin, M. & Çınar, M. (2017). *Ekonometrik Zaman Serileri Analizi: EViews Uygulamalı* (Fifth Edition). Bursa: Dora Yayıncılık.

Smith, G. P. (2012). Google Internet search activity and volatility prediction in the market for foreign currency. *Finance Research Letters*, *9*, 103–110.